

इंटरनेट

मानक

Disclosure to Promote the Right To Information

Whereas the Parliament of India has set out to provide a practical regime of right to information for citizens to secure access to information under the control of public authorities, in order to promote transparency and accountability in the working of every public authority, and whereas the attached publication of the Bureau of Indian Standards is of particular interest to the public, particularly disadvantaged communities and those engaged in the pursuit of education and knowledge, the attached public safety standard is made available to promote the timely dissemination of this information in an accurate manner to the public.

“जानने का अधिकार, जीने का अधिकार”

Mazdoor Kisan Shakti Sangathan

“The Right to Information, The Right to Live”

“पुराने को छोड़ नये के तरफ”

Jawaharlal Nehru

“Step Out From the Old to the New”

IS 7300 (2003): Methods of Regression and Correlation [MSD
3: Statistical Methods for Quality and Reliability]



“ज्ञान से एक नये भारत का निर्माण”

Satyanarayan Gangaram Pitroda

“Invent a New India Using Knowledge”



“ज्ञान एक ऐसा खजाना है जो कभी चुराया नहीं जा सकता है”

Bhartrhari—Nitiśatakam

“Knowledge is such a treasure which cannot be stolen”

BLANK PAGE



भारतीय मानक
समाश्रयण और सहसंबंध की पद्धतियाँ
(दूसरा पुनरीक्षण)

Indian Standard
METHODS OF REGRESSION AND CORRELATION
(*Second Revision*)

ICS 03.120.30

© BIS 2003

BUREAU OF INDIAN STANDARDS
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI 110002

FOREWORD

This Indian Standard (Second Revision) was adopted by the Bureau of Indian Standards, after the draft finalized by the Statistical Method for Quality and Reliability Sectional Committee had been approved by the Management and Systems Division Council.

The study of the relationship between two variables is of fundamental importance in industry. For example, in the building industry, while studying the properties of cement, it may be necessary to estimate the effect of curing time on the compressive strength. In such problems, where one variable is of particular interest for studying the effect of the other variable on it, the concept of regression is quite useful. The regression technique is also helpful for the purpose of prediction.

In some problems, the relationship between two variables may be of great interest, for example, in the case of steel, one can study tensile strength by using hardness test, as the latter has a strong relationship with the former. The determination of the extent of relationship between two variables leads to the concept of correlation.

This standard was originally published in 1974 to cover the statistical methods of regression and correlation in the case of two variables. This standard was revised in 1995 to include the concept of 'scatter diagram' more elaborately.

In view of the experience gained with the use of the standard in course of years, it was felt necessary to further revise it. In the revised version, following changes have been made:

- a) A table which gives the values of correlation coefficient (r) for different selected sample sizes has been included so that the sample correlation coefficient calculated value may directly be compared with this tabulated value to test whether the population correlation coefficient is zero or not,
- b) Confidence limits for the population regression line with example has been included,
- c) Many editorial corrections have been incorporated, and
- d) The concepts at many places have been elaborated for better understanding.

The composition of the Committee responsible for the formulation of this standard is given in Annex F.

Indian Standard

METHODS OF REGRESSION AND CORRELATION

(*Second Revision*)

1 SCOPE

This standard covers the statistical methods of linear regression and correlation in the case of two variables. The computations have been illustrated with examples.

2 REFERENCES

The following standards contain provisions, which through reference in this text constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below:

<i>IS No.</i>	<i>Title</i>
6200 (Part 1) : 1995	Statistical tests of significance : Part 1 <i>t</i> -Normal and <i>F</i> -tests (<i>second revision</i>)
7920 (Part 1) : 1994	Statistical vocabulary and symbols : Part 1 Probability and general statistical terms (<i>second revision</i>)
8900 : 1978	Criteria for the rejection of outlying observations
9300 (Part 1) : 1979	Statistical models for industrial applications: Part 1 Discrete models

3 TERMINOLOGY

For the purpose of this standard the definitions given in IS 7920 (Part 1) shall apply.

4 BASIC CONCEPTS

4.1 Scatter Diagram

4.1.1 The scatter diagram is useful to know the presence of the relationship or the nature of the relationship between two variables, if any. The relationship can be a cause and effect relationship, a relationship between one cause and the other, or a relationship between one effect and the other.

4.1.2 Scatter diagram can even be used by the operators to find the relationship between two variables, if any. This may lead to taking appropriate actions for quality improvement.

4.1.3 A scatter diagram is prepared by plotting the paired data in an *X-Y* plane. It is desirable to have more than 30 pairs of data. Of the two variables, one is said

to be independent and the other dependent and it is usual to regard the independent variable as *x* and the dependent variable as *y*. Since the range of data varies widely, the origin of zero is sometimes inconvenient to prepare a well-balanced scatter diagram. The data ranges are suitably presented on convenient scales so that spread is close to a square and large enough for individual perception.

4.1.4 The problem of outliers is encountered in the actual preparation of scatter diagrams. Outliers are too widely separated from the data set. If there are few outliers, they should be eliminated from the data. For guidance on the criteria for rejection of outliers, reference is invited to IS 8900. If there are many (generally more than 25 percent) outliers, the causes for the same should be investigated and corrective action taken. Thereafter, fresh data needs to be collected for plotting the scatter diagram.

4.1.5 Interpretation of a Scatter Diagram

When a scatter diagram is prepared, it is important to interpret it accurately and take necessary measures. For this purpose, the scatter diagram should be carefully observed for the relationship between two variables. The interpretation of the scatter diagrams is explained as follows:

- a) *Positive relationship* — In a scatter diagram, if *y* increases with increase in *x*, then the relationship is said to be positive. When the points are close to a straight line [see Fig. 1 (a)], the relationship is called a positive linear relationship. Under such conditions control on *y* (the dependent variable) can be achieved by exercising control on *x* (the independent variable).
- b) *Negative relationship* — In a scatter diagram, if *y* decreases with increase in *x*, then the relationship is said to be negative [see Fig. 1 (b)]. In this case, similar interpretation as given for (a) holds good.
- c) *Weak relationship* — Sometimes the relationships may not be as clearly evident as in (a) or (b) [see Fig. 1 (c)]. Further investigations may be required to find out the reasons, if any, for the wider scatter. Possibly one factor alone is not sufficient to explain the relationship fully or there could be wide

measurement errors. The relationship may not be useful for control purposes in such a situation.

- d) *No relationship* — In a scatter diagram [see Fig. 1 (d)], no relationship can be noted between x and y . If the presence of relationship is expected on technological considerations, the causes/effects may be examined from other viewpoints. In such a situation the possibility of stratifying the data may also be looked into [see 4.1.5 (e)].
- e) *Relationship revealed by stratification* — The scatter diagram [see Fig. 1 (e)], shows no relationship at a glance, but if the data is classified into some different groups a relationship may be possible. In this diagram,

the presence of relationship can be confirmed definitely by stratifying the data into three groups marked with: \bullet , Δ and χ .

- f) *Non-linear relationship* — In a scatter diagram there may be relationship between x and y but is non linear. For example, in Fig. 1 (f), y increases with an increase in x until a certain point, but decreases with an increase in x beyond that point. Such relationship is called non-linear relationship and can be treated otherwise. In such situation, it is convenient to locate optimal combination for x and y .
- g) *Insufficient data range* — When attention is paid only to the points marked with Δ , there seems to be no relationship between x and y ,

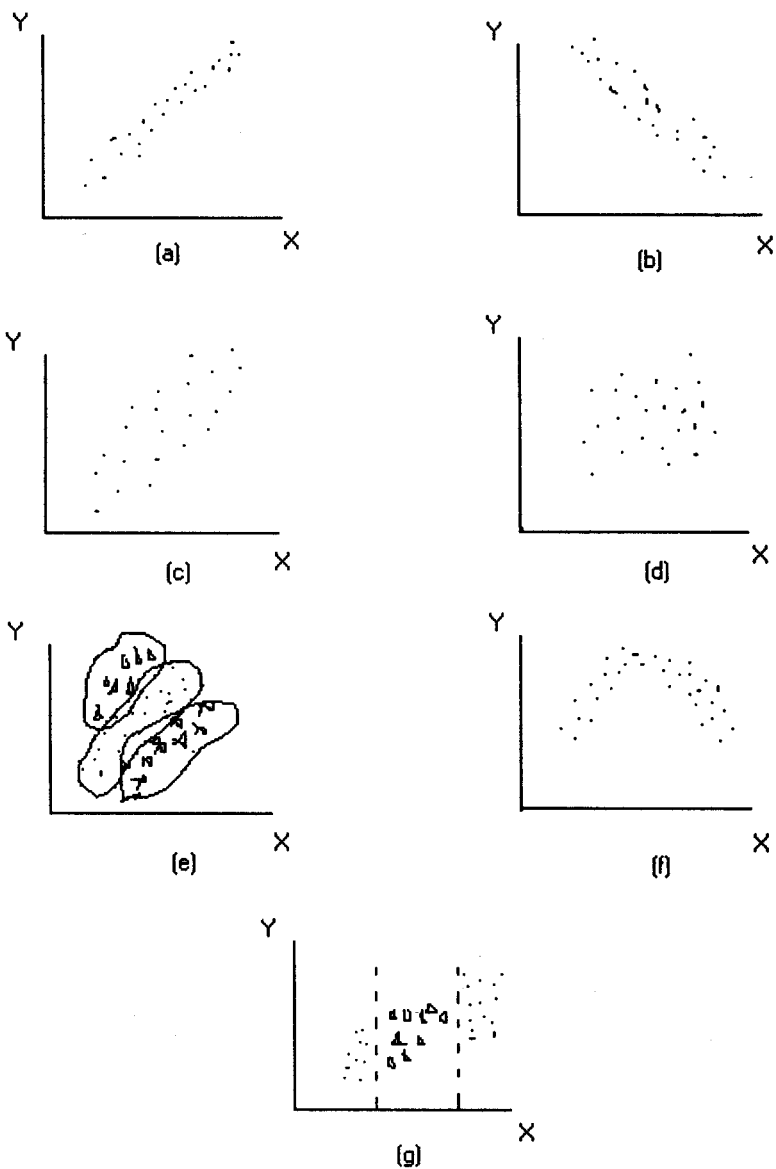


FIG. 1 VARIOUS SCATTER DIAGRAMS

as shown in Fig.1 (g), but positive linear relationship is noted when points are observed in a little wider range. Accordingly, it is necessary to examine carefully the appropriateness of the range of x even when no relationship is suggested in the diagram prepared for the first time.

4.2 Regression

Regression deals with situations when one variable is dependent on the other variable. For example, the two variables may be the quantities of the carbon steel and alloy steel produced from the same raw material or charge, elongation of boiler plate and the amount of tension applied, amount of rainfall and the yield of a crop, and so on. Of the two variables, one is independent (generally measurable) and the other is dependent (desired to be controlled). Thus, it is evident that the production of alloy steel depends on the production of carbon steel so that the quantity of carbon steel produced could be considered as the independent variable and that of alloy steel as the dependent variable.

4.3 Correlation

Correlation deals with the relationship between two factors or variables. The degree or intensity of the linear relationship is measured by correlation coefficient. It may be mentioned that in the study of correlation, it is not the intention to find the effect of one variable over the other as in the case of regression analysis but it is to find the degree to which the variables vary together owing to influences which affect both of them. However, the mere existence of high value of the correlation coefficient is not necessarily indicative of the underlying relationship between the two variables. Such a value can at times be purely accidental, the two variables having no connection whatsoever. In such cases, the correlation coefficient may be spurious.

4.4 Before carrying out any regression or correlation study, it is desirable to look at the scatter diagram to locate the outlier(s), if any and eliminate them.

5 REGRESSION ANALYSIS

5.1 Regression Coefficient

In a scatter diagram of type [see 4.1.5 (a) or (b)] a straight line could be fitted to the observed values which is of the form $y = a + bx$, where y is the dependent variable and x the independent variable. The quantity a in the above equation represents the value of y when $x = 0$, and b denotes the slope of the line and is known as the regression coefficient which may be negative or positive depending on the orientation of the line with respect to the axes. Physically, b indicates the rate of increase or decrease in the value of y for unit increase

or decrease in the value of x . The regression line is also used for prediction purposes. Normally, extrapolation is not recommended, and when necessary, it should be used cautiously.

5.1.1 The relationship of the type $y = a + bx$ encountered in the regression analysis is not generally reversible and is based on the status of the variables concerned. Therefore, this type of relationship should not be used for predicting x for given y . However, mathematically it is possible to find relationship of the type $x = a' + b'y$ and then the regression lines intersect at the point (x, y) in the x, y plane.

5.2 Method of Calculation (Ungrouped Data)

5.2.1 Let there be n pairs of observations for x and y corresponding to the items in the sample. For fitting the regression line the following expressions are then calculated:

- Average of x $\left(\bar{x} = \frac{\sum x}{n} \right)$
- Average of y $\left(\bar{y} = \frac{\sum y}{n} \right)$
- Corrected sum of squares for x

$$\sum (x - \bar{x})^2 = \sum x^2 - \left[(\sum x)^2 / n \right]$$
- Corrected sum of squares for y

$$\sum (y - \bar{y})^2 = \sum y^2 - \left[(\sum y)^2 / n \right]$$
- Corrected sum of products

$$\sum (x - \bar{x})(y - \bar{y}) = \sum xy - \left[(\sum x)(\sum y) / n \right]$$

NOTE — A suitable proforma as given in Annex A may be helpful in the above computations.

5.2.2 From the above quantities the regression coefficient b or b' is calculated as:

$$b = \frac{\text{Corrected sum of products}}{\text{Corrected sum of squares for } x}$$

$$b' = \frac{\text{Corrected sum of products}}{\text{Corrected sum of squares for } y}$$

Also the constant a or a' of the regression equation is obtained as:

$$a = \bar{y} - b\bar{x}$$

$$a' = \bar{x} - b'\bar{y}$$

5.2.3 When the regression model is not of the linear type and involves powers or exponentials, the model may be reduced to the linear type with the help of the logarithmic transformation. Thereafter, the fitting of the regression line is exactly similar to the one explained in 5.2.2.

5.2.4 Example

Table 1 gives the Brinell hardness number and the tensile strength (expressed in units of megapascals) for 15 specimens of cold drawn copper. Consider Brinell hardness number as the independent variable (*x*) and tensile strength as the dependent variable (*y*). It is intended to fit a regression line to the data.

5.2.4.1 Plotting the data given in Table 1 as a scatter diagram wherein the Brinell hardness number is measured along the *X*-axis and the tensile strength along the *Y*-axis, Fig. 2 is obtained, from which the linear trend of the points is self-evident. For the sake of better understanding, the regression line applicable to the data is also drawn in Fig. 2.

Table 1 Hardness and Tensile Strength Values of Cold Drawn Copper
(Clauses 5.2.4, 5.2.4.1 and 5.2.4.2)

Sl No.	Specimen No.	Brinell Hardness <i>x</i>	Tensile Strength <i>y</i>
(1)	(2)	(3)	(4)
i)	1	104.2	268.0
ii)	2	106.1	278.6
iii)	3	105.6	275.0
iv)	4	106.3	281.5
v)	5	101.7	232.4
vi)	6	104.4	272.2
vii)	7	102.0	227.5
viii)	8	103.8	255.1
ix)	9	104.0	259.5
x)	10	101.5	229.0
xi)	11	101.9	233.8
xii)	12	100.6	205.9
xiii)	13	104.9	272.0
xiv)	14	106.2	280.3
xv)	15	103.1	242.2

5.2.4.2 From the data in Table 1, various computations are obtained as follows:

$\Sigma x = 1\,556.3$ $\bar{x} = 103.75$

$\Sigma y = 3\,813.0$ $\bar{y} = 254.2$

$\Sigma x^2 = 161\,520.87$

$\Sigma xy = 396\,226.39$

Corrected sum of

= 161 520.87 – [(1 556.3)² /15]

squares for *x*

= 161 520.87 – 161 471.31

= 49.56

Corrected sum of

= 396 226.39 – $\left(\frac{1\,556.3 \times 3\,813}{15}\right)$

products

= 396 226.39 – 395 611.46

= 614.93

$b = 614.93/49.56 = 12.4$

$a = \bar{y} - b \bar{x} = 254.2 - 1\,286.5 = -1\,032.3$

Hence regression line is obtained as

$y = -1\,032.3 + 12.4\,x$

5.2.4.3 For simplifying the computational work involved in fitting a regression line, change of origin is often helpful in one or both the variables. Thus, for the example worked out in 5.2.4.2, if the variables

x and *y* are changed to *u* and *v* such that *u* = *x* – 100 and *v* = *y* – 250, then the computations would be as follows:

$\Sigma u = 56.3$ $\bar{u} = 3.75$

$\Sigma v = 63.0$ $\bar{v} = 4.20$

$\Sigma u^2 = 260.87$

$\Sigma uv = 851.39$

$\Sigma (u - \bar{u})^2 = \Sigma u^2 - [(\Sigma u)^2/n] = 260.87 - 211.31 = 49.56$

$\Sigma (u - \bar{u}) (v - \bar{v}) = \Sigma uv - [(\Sigma u) (\Sigma v)/n]$
 $= 851.39 - 236.46 = 614.93$

$b = 614.93/49.56 = 12.4$ and $\bar{v} - b \bar{u} = 4.2 - 46.5 = -42.3$

Hence the regression line is obtained as $\bar{v} = -42.3 + 12.4\, \bar{u}$ which when transformed to the original variables, comes out as:

$(y - 250) = -42.3 + 12.4 (x - 100)$
that is $y = -1\,032.3 + 12.4\,x$

NOTE — It would be of interest to observe that the regression coefficient *b* is not affected by the change of origin of either or both the variables.

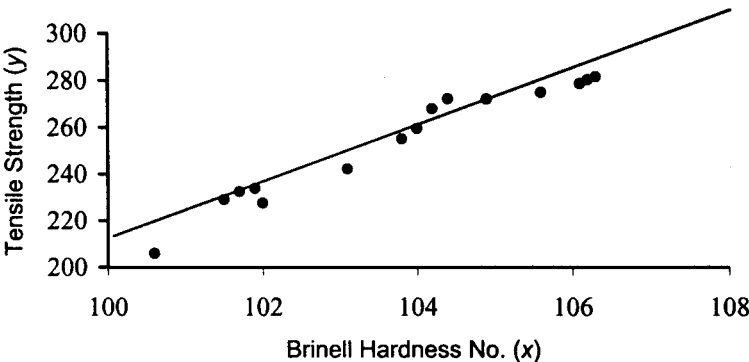


FIG. 2 SCATTER DIAGRAM ALONGWITH THE REGRESSION LINE

5.2.4.4 From this equation the expected value of tensile strength for any given Brinell hardness number could be obtained. Thus, when the hardness number is known as 105 the corresponding expected value of tensile strength would be 269.7 megapascals.

5.2.5 Construction of Confidence Limits for the Regression Line

The model is $y = \alpha + \beta x + \text{error}$

The estimates b of β and a of α are obtained for the example as :

$a = -1\,032.3, \qquad b = 12.4$

The error sum of squares ($\sigma^2_{y,x}$) is given by :

$\Sigma(y_i - a - b x_i)^2 / (15 - 2) = 30.813$

For a particular value of $X = x$ (Brinell hardness = x), the predicted value of the tensile strength (\hat{y}) is:

$\hat{y} = -1\,032.3 + 12.4 x$

The standard deviation of \hat{y} given $X = x$ is

$s(\hat{y}) = \sigma_{y,x} [(1/n) + \{(x - \bar{x})^2 / \Sigma(x - \hat{x})^2\}]^{1/2}$

Therefore, for a given x the confidence limits on the value of y are

$\hat{y} \pm t s(\hat{y}/x)$

where t is the value of a t distribution with $(n - 2 = 13)$ degrees of freedom.

Since we are interested in the confidence limits for the whole of the regression line, these limits for individual \hat{y} have to be relaxed. The appropriate multiplier is $(2 F)^{1/2}$ where F is the upper 5 percent tail of F distribution with degrees of freedom $\gamma_1 = 2$ and $\gamma_2 = (n - 2) = (15 - 2) = 13$. From the tables of F , the value of $F(2, 13)$, at 5 percent level of significance is 3.80. So the multiplier is $(2 \times 3.8)^{1/2} = 2.76$.

So the confidence limits for the regression line are given by:

$a + bx + 2.76 \left[\sqrt{30.813} \left\{ (1/15) + (x - \bar{x})^2 / 49.56 \right\} \right],$
and
 $a + bx - 2.76 \left[\sqrt{30.813} \left\{ (1/15) + (x - \bar{x})^2 / 49.56 \right\} \right]$

Therefore, the confidence limits of regression line for the data given in Table 1 have been calculated from the above expressions and are given in Table 2.

NOTE — The upper and lower limits for the regression line form a hyperbolic curve. When x is close to \bar{x} the contribution of this term is small. As x deviates from \bar{x} , the contribution of this term increases.

5.3 Method of Calculation (Grouped Data)

5.3.1 Sometimes, the observations on the two variables

x and y are presented in the form of a frequency table. In such situations the range of each variate is divided into a number of class intervals of equal width (say l_x for p classes of independent variable x and l_y for q classes of dependent variable y). The class width for x and y need not be equal, and the frequency f_{xij} in the cell is determined by the i th class interval of the first variate and j th class interval of the second variate. This would result in a bivariate frequency distribution table (see Annex B)

Table 2 Confidence Limits for Regression Line
(Clause 5.2.5)

x (1)	y (2)	Upper Limit (3)	Lower Limit (4)
100.6	215.14	223.05	207.23
101.5	226.30	232.59	220.01
102.0	232.50	237.99	227.01
103.1	246.14	250.36	241.91
103.8	254.82	258.78	250.85
104.2	259.78	263.86	255.70
104.9	268.46	273.14	263.78
105.6	277.14	282.78	271.50
106.3	285.82	292.63	279.01

5.3.2 As a first step for calculating the regression line, another proforma (see Annex C) is to be prepared.

5.3.3 The different entries in the above proforma are explained below:

- a) In the top row are given the mid-values of the class intervals for the independent variable x whereas in the first column are given mid-values of the class intervals for the dependent variable y . In the column f_y are given the total frequencies of the corresponding rows whereas in the row corresponding to f_x are given the total of the corresponding frequencies in the various columns.
- b) In the row corresponding to u are given the transformed variables for x which are obtained by subtracting an arbitrary quantity x_0 (preferably value of x closest to median) from each of the mid-values of the class intervals for x variate and dividing these differences by the width of the class intervals for x variate. That is, $u = (x - x_0)/l_x$, where l_x is the width of the class interval for x . A similar transformed variable v is given for the variate y in the respective column $v = (y - y_0)/l_y$.
- c) The next two rows, namely, uf_x and $u^2 f_x$ are self-explanatory. So also the two columns corresponding to vf_y and $v^2 f_y$.
- d) The row corresponding to V is obtained as sum of the products of v and the corresponding frequency in the column. So

also the column corresponding to U consists of entries obtained as the sum of products u with the corresponding frequency in the row.

- e) The last row uV as also the last column vU are self explanatory.

5.3.4 To ensure the correctness of computations it would be necessary to verify the following checks from the above proforma:

- The total frequency of the row corresponding to f_x should be equal to the total frequency of the column corresponding to f_y .
- The total of the row corresponding to V should be equal to the total of the column corresponding to $v f_y$.
- The total of the column corresponding to U should be equal to the total of the row corresponding to $u f_x$.
- The sum of the last row corresponding to uV should be equal to the sum of the last column corresponding to vU .

5.3.5 After the computations using the above proforma, the regression coefficient is calculated by the following formula:

$$b = [\Sigma uV - \{(\Sigma U)(\Sigma V)/n\}] l_y / [\Sigma u^2 f_x - (\Sigma u f_x)^2 / n] l_x$$

The constant of the regression line is obtained as:

$$a = y - b x = \{y_c + l_y (\Sigma v f_y / n)\} - b \{x_0 + l_x (\Sigma u f_x / n)\}$$

5.3.6 Example

Table 3 gives the distribution of 82 small and medium size sugar factories by the quantity of cane crushed (x) and the quantity of sugar produced (y). Fit a regression line.

As a first step, the computation are made in Table 4.

The regression coefficient is computed as:

$$b = [329 - \{(-78)(-35)/82\}] \times 2 / [410 - \{(-78)^2/82\}] \times 20 = 0.0881 \approx 0.09$$

The constant of the regression line is obtained as:

$$\begin{aligned} a &= \{12 - (2 \times 35/82)\} - 0.0881 \{140 - (20 \times 78/82)\} \\ &= 11.15 - 0.0881 \times 120.98 = 11.15 - 10.66 \\ &= 0.49 \end{aligned}$$

Hence the regression line is obtained as:

$$y = 0.49 + 0.09 x$$

This line can be used for predicting the quantity of sugar produced knowing the quantity of the cane crushed. Thus if 100 tonne of cane is crushed, then by the above equation, 9.5 tonne of sugar can be expected to be produced.

NOTE — For the purpose of prediction, the regression line may be used only within the range of the independent variable and in the vicinity of the terminal values.

5.4 Testing for Regression Coefficient

5.4.1 From the way the regression coefficient is to be calculated, it is obvious that its value depends on the sample observations. Hence if a new set of observations is obtained from the same population and the corresponding regression coefficient is calculated, it may not necessarily be the same as earlier one. Because of this fluctuation it is necessary to test whether the regression coefficient as calculated from the sample observations differs significantly from some specified value which may correspond to the entire population. Sometimes, the specified value may also be a rounded off value which seems more feasible as the population regression coefficient. However, for any testing of the regression coefficient to be valid, it is assumed that both the independent and the dependent variables

Table 3 Frequency Distribution of Sugar Factories by Cane Crushed and Sugar Produced
(Clause 5.3.6)

Sl No.	Sugar Produced in 1 000 tonne (y)	Cane Crushed in 1 000 tonne (x)									
		30-49	50-69	70-89	90-109	110-129	130-149	150-169	170-189	190-209	210-229
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
i)	3.0- 4.9	2	—	—	—	—	—	—	—	—	—
ii)	5.0- 6.9	—	6	2	1	—	—	—	—	—	—
iii)	7.0- 8.9	—	—	8	4	1	—	—	—	—	—
iv)	9.0-10.9	—	—	—	11	6	1	—	—	—	—
v)	11.0-12.9	—	—	—	1	7	7	2	1	—	—
vi)	13.0-14.9	—	—	—	—	—	9	4	—	—	—
vii)	15.0-16.9	—	—	—	—	—	—	1	—	1	—
viii)	17.0-18.9	—	—	—	—	—	—	—	1	2	—
ix)	19.0-20.9	—	—	—	—	—	—	—	—	1	1
x)	21.0-22.9	—	—	—	—	—	—	—	—	1	1

Table 4 Proforma for Computation of Regression Line for the Data Given in Table 3

(Clause 5.3.6)

$y \backslash x$	40	60	80	100	120	140	160	180	200	220	f_x	v	vf_y	v^2f_y	U	vU
4.0	2	—	—	—	—	—	—	—	—	—	2	-4	-8	32	-10	40
6.0	—	6	2	1	—	—	—	—	—	—	9	-3	-27	81	-32	96
8.0	—	—	8	4	1	—	—	—	—	—	13	-2	-26	52	-33	66
10.0	—	—	—	11	6	1	—	—	—	—	18	-1	-18	18	-28	28
12.0	—	—	—	1	7	7	2	1	—	—	18	0	0	0	-5	0
14.0	—	—	—	—	—	9	4	—	—	—	13	1	13	13	4	4
16.0	—	—	—	—	—	—	1	—	1	—	2	2	4	8	4	8
18.0	—	—	—	—	—	—	—	1	2	—	3	3	9	27	8	24
20.0	—	—	—	—	—	—	—	—	1	1	2	4	8	32	7	28
22.0	—	—	—	—	—	—	—	—	1	1	2	5	10	50	7	35
f_y	2	6	10	17	14	17	7	2	5	2	82	5	-35	313	-78	329
u	-5	-4	-3	-2	-1	0	1	2	3	4						
uf_y	-10	-24	-30	-34	-14	0	7	4	15	8	-78					
u^2f_y	50	96	90	68	14	0	7	8	45	32	410					
v	-8	-18	-22	-22	-8	8	6	3	17	9	-35					
uV	40	72	66	44	8	0	6	6	51	36	329					

follow a normal distribution. For further details of the normal distribution, see IS 9300 (Part 1).

5.4.2 Ungrouped Data

The value of the regression coefficient as calculated from the sample is an estimate of the true regression coefficient for the entire population. To judge whether the population regression coefficient differs significantly from a specified value, β_0 , the null hypothesis, $H_0: \beta = \beta_0$ is tested against the alternative hypothesis, $H_1: \beta \neq \beta_0$ by computing the following test-statistic:

$$t = \frac{(b - \beta_0) \left\{ \sum (x - \bar{x})^2 \right\}^{1/2}}{\left[\left\{ \sum (y - \bar{y})^2 - b \sum (x - \bar{x})(y - \bar{y}) \right\} / (n - 2) \right]^{1/2}}$$

b = regression coefficient computed from the data,

β_0 = specified value of the regression coefficient,

$\sum (x - \bar{x})^2$ = corrected sum of squares for x ,

$\sum (y - \bar{y})^2$ = corrected sum of squares for y ,

$\sum (x - \bar{x})(y - \bar{y})$ = corrected sum of products, and

n = sample size.

The calculated value of t shall be compared with the tabulated value of t [see Annex B of IS 6200 (Part 1)] at desired level of significance (normally 5 percent) and for $(n - 2)$ degrees of freedom. If the calculated value of t is greater than or equal to the tabulated value, the null hypothesis is rejected and the alternative hypothesis that the population regression coefficient is significantly different from the specified value of β_0 is accepted, otherwise not.

As a particular case, when $\beta_0 = 0$ the test would be

used to verify the assumption that the change in the independent variable does not affect the dependent variable in the population in a systematic manner.

5.4.2.1 Example

In the illustration given in 5.2.4 concerning the regression equation for predicting the tensile strength from Brinell hardness number for cold drawn copper, it may be of interest to test whether the population regression coefficient is significantly different from the specified value of 11.0 which was found to hold good in an earlier investigation done on a large number of samples.

In this case, $H_0: \beta = 11.0$ and $H_1: \beta \neq 11.0$

The t -statistic is calculated as follows:

$$\begin{aligned} t &= (12.4 - 11.0) (49.56)^{1/2} / [\{ 8 \ 030.9 - (12.4 \times 614.93) \} / 13]^{1/2} \\ &= 1.4 \times 7.04 / 5.59 = 1.76 \end{aligned}$$

The tabulated value of t at 5 percent level of significance and for 13 degrees of freedom is given as 2.160. Since the calculated value is less than 2.160, H_0 is not rejected and it is concluded that the population regression coefficient is not significantly different from 11.0.

5.4.3 Grouped Data

In the case of grouped data for testing whether the population regression coefficient is significantly different from the specified value of β_0 , the null hypothesis, $H_0: \beta = \beta_0$ is tested against the alternative hypothesis, $H_1: \beta \neq \beta_0$ and the t -statistic is computed as:

$$\begin{aligned} t &= A / \sqrt{B} \\ A &= (b - \beta_0) [\sum u^2 f_x - \{ (\sum u f_x)^2 / n \}]^{1/2} \cdot l_x \end{aligned}$$

$$B = \frac{\left[\sum v^2 f_y - \frac{(\sum v f_y)^2}{n} \right] l_y^2 - b \left[\sum u v - \frac{(\sum U)(\sum V)}{n} \right] l_x l_y}{n - 2}$$

The testing can be done on the same lines as indicated in 5.4.2.

5.4.3.1 In the example given in 5.3.6 for predicting the quantity of sugar produced from the quantity of cane crushed, it may be of interest to examine whether the population regression coefficient is significantly different from zero.

In this case, $H_0 : \beta = 0; H_1 : \beta \neq 0$

t -statistic is computed as follows:

$$A = 0.09 [410 - \{(-78)^2/82\}]^{1/2} \times 20 = 32.9850$$

$$B = \left[\left\{ 313 - (-35)^2/82 \right\} \times 4 \right] - \frac{[0.09 \{ 329 - (-78)(-35)/82 \} \times 20 \times 2]}{80} = 1.5962$$

$$t = 32.9850 / \sqrt{1.5962} = 32.9850 / 1.2634 = 26.1$$

Since the tabulated value of t at 5 percent level of significance and 80 degrees of freedom is given as 1.96, H_0 is rejected and it is concluded that the population regression coefficient is significantly different from zero.

6 CORRELATION

6.1 Correlation Coefficient

6.1.1 The correlation coefficient is usually denoted by the symbol ρ with respect to the population under study. When the study is based on a sample drawn from a population, it is denoted by the symbol r . Values of the correlation coefficient lie between -1 and $+1$. If it is $+1$, perfect positive correlation exists between the

two factors under study which implies that a definite increase in one factor is accompanied by a proportionate increase in the other factor.

6.1.2 If $r = -1$ then perfect negative correlation is present, meaning thereby that a definite increase in one factor is followed by proportionate decrease in other factor, or *vice versa*. If the correlation coefficient is zero then the two factors are said to be uncorrelated. The correlation coefficient is a pure number and its magnitude is unaffected by the scale in which the two variables x and y are measured.

6.1.3 Figure 3 gives the scatter diagram for the two variables x and y in three situations, namely, when the correlation coefficient is high positive (say, 0.9), zero and high negative (say, -0.9).

6.2 Method of Calculation (Ungrouped Data)

6.2.1 Let there be n paired observations x and y corresponding to the items in the sample. The average of x , average of y , corrected sum of squares for x , corrected sum of squares for y and corrected sum of products are then calculated as given in 5.2.1.

6.2.2 From the above quantities the correlation coefficient is calculated as follows:

$$r = \frac{\text{Corrected sum of products}}{\left[\frac{(\text{Corrected sum of squares for } x) \times (\text{Corrected sum of squares for } y)}{} \right]^{1/2}}$$

6.2.3 When the status of the two variables are not known (see 5.1.1) the two regression coefficients obtained in fitting the lines $y = a + b x$ and $x = a' + b' y$ and the correlation coefficient r are related as $r = b b'$.

6.2.4 Example

An investigation was carried out on 4-litre paint tins for finding the correlation between the capacity as calculated from the base dimensions and height and

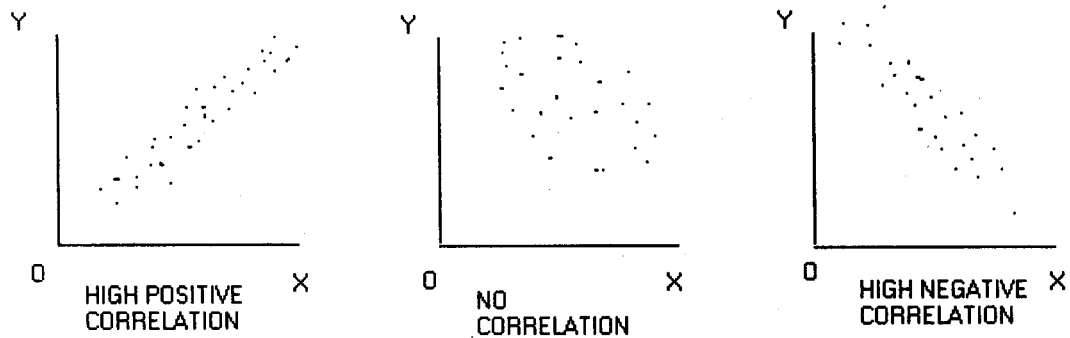


FIG. 3 SCATTER DIAGRAM

the actual measured capacity with a view to reducing the testing for the latter characteristic which was more time consuming as compared to dimensional checking. Table 5 gives the data obtained on 35 such tins. From the data tabulated in Table 5 the various computations are obtained as follows:

$$\begin{aligned}\Sigma x &= 165.930 & \bar{x} &= 4.74 \\ \Sigma y &= 158.095 & \bar{y} &= 4.517 \\ \Sigma x^2 &= 786.678\ 154 \\ \Sigma y^2 &= 714.131\ 775 \\ \Sigma xy &= 749.518\ 555 \\ \text{Corrected sum of squares for } x &= 0.027\ 728 \\ \text{Corrected sum of squares for } y &= 0.016\ 660 \\ \text{Corrected sum of products} &= 0.012\ 745\end{aligned}$$

Hence the correlation coefficient is equal to

$$r = 0.012\ 745 / (0.027\ 728 \times 0.016\ 660)^{1/2} = 0.59$$

NOTE — It may be of interest to observe that the correlation coefficient r is not affected by the change of origin and scale for either or both the variables. Hence the computations given in the above examples can be simplified considerably by making the transformations as follows:

$$\begin{aligned}u &= (x - 4.700) \times 1\ 000 \\ v &= (y - 4.500) \times 1\ 000\end{aligned}$$

6.3 Method of Calculation (Grouped Data)

6.3.1 If the observations on the two variables x and y are presented in the form of a frequency table in which the range of each variate is divided into a number of class intervals and the frequency f_{xij} corresponds to the cell determined by the i th class interval of the first variate and the j th class interval of the second variate then the initial computations for obtaining the correlation coefficient would be exactly the same as given in 5.3.1 and 5.3.2.

6.3.2 After the necessary tabulation of initial computations (see 5.3.2) are made, the correlation coefficient is obtained by the following formula:

$$r = \frac{\Sigma uV - \{(\Sigma U)(\Sigma V)/n\}}{\left[\left\{ \Sigma u^2 f_x - (\Sigma uf_x)^2 / n \right\} \left\{ \Sigma v^2 f_y - (\Sigma vf_y)^2 / n \right\} \right]^{1/2}}$$

6.3.3 Example

Table 6 gives the distribution of 100 casts of steel by the percentage of iron in the form of pig iron (x) and the lime consumption in quintal per cast (y). As a first step the computations as in Table 7 are made:

The correlation coefficient r is calculated as:

$$r = \frac{16 - 76(-88)/100}{\left[\left\{ 254 - (76)^2 / 100 \right\} \left\{ 264 - (-88)^2 / 100 \right\} \right]^{1/2}}$$

Table 5 Capacities of 4-Litre Paint Tins

(Clause 6.2.4)

SI No. of Tin	Calculated Capacity x	Measured Capacity y
(1)	(2)	(3)
1	4.732	4.530
2	4.735	4.540
3	4.756	4.550
4	4.709	4.540
5	4.708	4.540
6	4.768	4.500
7	4.726	4.490
8	4.744	4.510
9	4.686	4.485
10	4.693	4.495
11	4.695	4.480
12	4.694	4.485
13	4.692	4.485
14	4.727	4.490
15	4.729	4.490
16	4.745	4.500
17	4.741	4.500
18	4.704	4.510
19	4.741	4.510
20	4.745	4.515
21	4.771	4.520
22	4.774	4.515
23	4.768	4.510
24	4.758	4.525
25	4.772	4.520
26	4.779	4.550
27	4.763	4.550
28	4.757	4.540
29	4.781	4.550
30	4.784	4.555
31	4.758	4.515
32	4.753	4.520
33	4.753	4.525
34	4.732	4.525
35	4.757	4.530

$$= \frac{16 + 66.88}{\left[(254 - 57.76)(264 - 77.44) \right]} = 0.43$$

6.4 Testing for Correlation Coefficient

6.4.1 Correlation coefficient as calculated from the sample data is the estimate of the correlation coefficient applicable to all the items in the population. It may, however, sometimes be necessary to test whether the population correlation coefficient differs significantly from a specified value of ρ_0 . The corresponding tests to be performed when ρ is equal to zero and when ρ_0 is a non-zero value are slightly different and are given below.

6.4.2 To judge whether the population correlation coefficient differs significantly from zero (that is $\rho = 0$), the null hypothesis, $H_0 : \rho = 0$ is tested against the alternative hypothesis, $H_1 : \rho \neq 0$ by computing the

Table 6 Frequency Distribution of Percentage of Pig Iron (x) and Lime Consumption (y)
(Clause 6.3.3)

Sl No.	Lime Consumption in Quintal per Cast (y)	Percentage of Pig Iron (x)						
		20-24	25-29	30-34	35-39	40-44	45-49	50-54
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
i)	100-124	—	—	—	1	—	—	—
ii)	125-149	—	2	1	7	1	—	—
iii)	150-174	1	1	6	6	1	2	2
iv)	175-199	—	—	6	12	3	6	5
v)	200-224	—	—	—	3	8	11	3
vi)	225-2	49	—	—	1	1	2	3
vii)	250-274	—	—	—	—	—	1	1
viii)	275-299	—	—	—	1	—	—	—
ix)	300-324	—	—	—	—	—	1	—

Table 7 Computations
(Clause 6.3.3)

$y \backslash x$	22	27	32	37	42	47	52	f_x	v	vf_y	v^2f_y	U	vU
112	—	—	—	1	—	—	—	1	—4	—4	16	0	0
137	—	2	1	7	1	—	—	11	—3	—33	99	—4	12
162	1	1	6	6	1	2	2	19	—2	—38	76	0	0
187	—	—	6	12	3	6	5	32	—1	—32	32	24	—24
212	—	—	—	3	8	11	3	25	0	0	0	39	0
237	—	—	1	1	2	3	1	8	1	8	8	10	10
262	—	—	—	—	—	1	1	2	2	4	8	5	10
287	—	—	—	1	—	—	—	1	3	3	9	0	0
312	—	—	—	—	—	1	—	1	4	4	16	2	8
f_y	1	3	14	31	15	24	12	100	—	—88	264	76	16
u	—3	—2	—1	0	1	2	3	—	—	—	—	—	—
uf_y	—3	—6	—14	0	15	48	36	76	—	—	—	—	—
u^2f_y	9	12	14	0	15	96	108	254	—	—	—	—	—
V	—2	—8	—20	—45	—6	—1	—6	—88	—	—	—	—	—
uV	6	16	20	0	—6	—2	—18	16	—	—	—	—	—

following statistic:

$$t = r (n - 2)^{1/2} / (1 - r^2)^{1/2}$$

where r = correlation coefficient as computed from the sample and n is the sample size.

The value of t so calculated shall be compared with the tabulated value of t [see Annex B of IS 6200 (Part 1)] at the desired level of significance (normally 5 percent) and for $(n - 2)$ degrees of freedom.

If the calculated value of t is greater than or equal to the tabulated value, H_0 is rejected and the population correlation coefficient is said to be significantly different from zero, meaning thereby, that the two factors under consideration are correlated. However, if the calculated value of t is less than the tabulated value, H_0 is not rejected and it indicates that the sample data does not show any evidence that the factors under consideration are correlated.

For some selected values of sample sizes n , the table values of r have been calculated for critical values of t at 5 percent and 1 percent level of significance and

given in Annex D. If the calculated value of correlation coefficient value is less than the tabulated value, the null hypothesis is accepted, otherwise not.

6.4.2.1 Example

In the illustration given in 6.3.3 wherein the correlation coefficient between the percentage of pig iron and lime consumption in quintal per cast was computed as 0.43, if it is intended to test whether the population correlation coefficient is significantly different from zero, the null hypothesis is $H_0: \rho = 0.43$ and the alternative hypothesis is $H_1: \rho \neq 0.43$. The t -statistic is computed as:

$$t = r (n - 2)^{1/2} / (1 - r^2)^{1/2}$$
$$t = (0.43) \sqrt{98} / (0.815)^{1/2} = 4.71$$

Since the tabulated value of t distribution with 98 degrees of freedom and 5 percent level of significance is near about 1.96, H_0 is rejected and it is concluded that the population correlation coefficient is significantly different from zero, that is, the variables are associated to a significant extent.

6.4.3 For judging whether the population correlation coefficient differs significantly from the specified value ρ_0 (other than zero), the null hypothesis, $H_0 : \rho \neq \rho_0$ shall be tested against the alternative hypothesis, $H_1 : \rho \neq \rho_0$. The sample correlation coefficient r and the specified value ρ_0 shall be transformed into z and z_0 with the help of Annex E and the following statistic shall be computed:

$$t = |z - z_0| (n - 3)^{1/2}$$

where $|z - z_0|$ denotes the value of the difference between z and z_0 ignoring the sign.

If the value of this *statistic* is less than or equal to 1.96 (corresponding to 5 percent level of significance of the normal deviate), then H_0 is not rejected and it indicates that the population correlation coefficient is not significantly different from ρ_0 . In case the calculated value of the normal deviate is more than 1.96, H_0 is rejected and it indicates that the population correlation coefficient is significantly different from the specified value of ρ_0 .

NOTE — If the level of significance chosen is 1 percent then

instead of 1.96 the value 2.58 is to be used in the above comparison.

6.4.3.1 Example

In the illustration given under 6.2.4 wherein the correlation coefficient between the calculated capacity and measured capacity of 4-litre paint tins was computed as 0.59, it may be of interest to test whether the population correlation coefficient is significantly different from 0.70.

In this case, null hypothesis is $H_0 : \rho = \rho_0$ and the alternative hypothesis is $H_1 : \rho \neq \rho_0$.

From Annex E, the value of z corresponding to $r = 0.59$ is obtained as 0.677 7 and that of z_0 corresponding to $\rho_0 = 0.70$ is obtained as 0.867 3.

Hence $|z - z_0| (n - 3)^{1/2} = |0.677 7 - 0.867 3| \sqrt{32} = 0.189 6 \times 5.66 = 1.073$

Since this value is less than 1.96, H_0 is not rejected and there is not enough evidence to conclude that population correlation coefficient is significantly different from 0.70.

ANNEX A

(Clause 5.2.1)

PROFORMA FOR COMPUTATION OF CORRELATION/REGRESSION

Product:

Independent variable (x):

Unit of measurement:

Dependent variable (y):

Unit of measurement:

Sl No.	x	y	$u = \frac{x - x_0}{l_x}$	$v = \frac{y - y_0}{l_y}$	u^2	v^2	uv
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

Total

Mean

NOTE — In case the variables x and y are not transformed to u and v respectively, col 6, 7 and 8 may be utilized for tabulating x^2 , y^2 and xy .

ANNEX B

(Clause 5.3.1)

BIVARIATE FREQUENCY DISTRIBUTION TABLE

<div><div><div><div></div><div>x_i</div></div><div>y_j</div></div></div> <div><div><div><div></div><div>y_j</div></div><div>x_i</div></div></div>

f_{xij} = frequency in the (*i*, *j*) cell.

n = total frequency

NOTE — x_i is the mid-point of the interval *i*th column and y_j is the mid-point of interval *j*th row.

ANNEX C

(Clause 5.3.2)

PROFORMA FOR CALCULATING REGRESSION LINE FOR GROUPED DATA

<div><div><div><div></div><div>x</div></div><div>y</div></div></div>
--

ANNEX D*(Clause 6.4.2)***TABULATED VALUES OF r FOR 5 PERCENT AND 1 PERCENT LEVEL OF SIGNIFICANCE**

n	<i>Calculated Values of r</i>	
	5 percent Level of Significance	1 percent Level of Significance
3	1.000 0	1.000 0
4	1.000 0	1.000 0
5	0.954 0	0.985 7
6	0.891 0	0.955 9
7	0.829 4	0.918 8
8	0.774 3	0.880 1
9	0.726 6	0.842 6
10	0.685 5	0.807 6
11	0.649 8	0.775 5
12	0.618 8	0.746 1
13	0.591 5	0.719 3
14	0.567 4	0.694 8
15	0.545 7	0.672 3
16	0.526 5	0.651 8
17	0.508 8	0.632 9
18	0.493 0	0.615 4
19	0.478 4	0.599 1
20	0.465 0	0.584 0
21	0.452 6	0.570 1
22	0.441 2	0.556 9
23	0.430 7	0.544 7
24	0.420 7	0.533 2
25	0.411 5	0.522 3
26	0.402 8	0.512 2
27	0.394 7	0.502 4
28	0.387 0	0.493 4
29	0.379 7	0.484 7
30	0.372 7	0.476 4

ANNEX E*(Clause 6.4.3)***THE Z-TRANSFORMATION OF THE CORRELATION COEFFICIENT ($z = \tanh^{-1}r$)**

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.000 0	0.010 0	0.020 0	0.030 0	0.040 0	0.050 0	0.060 1	0.070 1	0.080 2	0.090 2
0.1	0.100 3	0.110 4	0.120 6	0.130 7	0.140 9	0.151 1	0.161 4	0.171 7	0.182 0	0.192 3
0.2	0.202 7	0.213 2	0.223 7	0.234 2	0.244 8	0.255 4	0.266 1	0.276 9	0.287 7	0.298 6
0.3	0.309 5	0.320 5	0.331 6	0.342 8	0.354 1	0.365 4	0.376 9	0.388 4	0.400 1	0.411 8
0.4	0.423 6	0.435 6	0.447 7	0.459 9	0.472 2	0.484 7	0.497 3	0.510 1	0.523 0	0.536 1
0.5	0.549 3	0.562 7	0.576 3	0.590 1	0.604 2	0.618 4	0.632 8	0.647 5	0.662 5	0.677 7
0.6	0.693 1	0.708 9	0.725 0	0.741 4	0.758 2	0.775 3	0.792 8	0.810 7	0.829 1	0.848 0
0.7	0.867 3	0.887 2	0.907 6	0.928 7	0.950 5	0.973	0.996	1.020	1.045	1.071
0.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422
0.9	1.472	1.528	1.589	1.658	1.738	1.832	1.946	2.092	2.298	2.647

ANNEX F

(Foreword)

COMMITTEE COMPOSITION

Statistical Methods for Quality and Reliability Sectional Committee, MSD 3

<i>Organization</i>	<i>Representative(s)</i>
Kolkata University, Kolkata	PROF S. P. MUKHERJEE (<i>Chairman</i>)
Bharat Heavy Electricals Limited, Hyderabad	SHRI S. N. JHA
	SHRI A. V. KRISHNAN (<i>Alternate</i>)
Continental Devices India Ltd, New Delhi	DR NAVIN KAPUR
	SHRI VIPUL GUPTA (<i>Alternate</i>)
Directorate General of Quality Assurance, New Delhi	SHRI S. K. SRIVASTVA
	LT-COL P. VIJAYAN (<i>Alternate</i>)
Laser Science and Technology Centre, DRDO, New Delhi	DR ASHOK KUMAR
Escorts Limited, Faridabad	SHRI C. S. V. NARENDRA
HMT Ltd, R & D Centre, Bangalore	SHRI K. VIJAYAMMA
Indian Agricultural Statistics Research Institute (IASRI), New Delhi	DR S. D. SHARMA
	DR A. K. SRIVASTAVA (<i>Alternate</i>)
Indian Association for Productivity, Quality & Reliability (IAPQR), Kolkata	DR B. DAS
Indian Institute of Management (IIM), Lucknow	PROF S. CHAKRABORTY
Indian Statistical Institute (ISI), Kolkata	PROF S. R. MOHAN
	PROF ARVIND SETH (<i>Alternate</i>)
National Institution for Quality and Reliability (NIQR), New Delhi	SHRI Y. K. BHAT
	SHRI G. W. DATEY (<i>Alternate</i>)
Powergrid Corporation of India Ltd, New Delhi	DR S. K. AGARWAL
	SHRI D. CHAKRABORTY (<i>Alternate</i>)
SRF Limited, Chennai	SHRI A. SANJEEVA RAO
	SHRI C. DESIGAN (<i>Alternate</i>)
Standardization, Testing and Quality Certification Directorate (STQCD), New Delhi	SHRI S. K. KIMOTHI
Tata Engineering and Locomotive Co Ltd (TELCO), Jamshedpur	SHRI P. N. SRIKANTH (<i>Alternate</i>)
	SHRI S. KUMAR
	SHRI SHANTI SARUP (<i>Alternate</i>)
University of Delhi, Delhi	PROF M. C. AGRAWAL
In personal capacity (B-109, Malviya Nagar, New Delhi 110017)	PROF A. N. NANKANA
In personal capacity (20/1, Krishna Nagar, Safdarjang Enclave, New Delhi 110029)	SHRI D. R. SEN
BIS Directorate General	SHRI P. K. GAMBHIR, Director & Head (MSD) [Representing Director General (<i>Ex-officio</i>)]
<i>Member Secretary</i>	
SHRI LALIT KUMAR MEHTA	
Deputy Director (MSD), BIS	

Basic Statistical Methods Subcommittee, MSD 3 : 1

Kolkata University, Kolkata	PROF S. P. MUKHERJEE (<i>Convener</i>)
Laser Science and Technology Centre, DRDO, New Delhi	DR ASHOK KUMAR
Indian Agricultural Statistics Research Institute (IASRI), New Delhi	DR S. D. SHARMA
	DR DEBABRATA RAY (<i>Alternate</i>)
Indian Association for Productivity, Quality and Reliability (IAPQR), Kolkata	DR B. DAS
Indian Institute of Management (IIM), Lucknow	DR A. LAHIRI (<i>Alternate</i>)
	PROF S. CHAKRABORTY

(Continued on page 15)

(Continued from page 14)

<i>Organization</i>	<i>Representative(s)</i>
Indian Statistical Institute (ISI), Kolkata	PROF S. R. MOHAN
National Institution for Quality and Reliability (NIQR), New Delhi	SHRI Y. K. BHAT SHRI G. W. DATEY (<i>Alternate</i>)
Powergrid Corporation of India Ltd, New Delhi	DR S. K. AGARWAL
Standardization, Testing and Quality Certification (STQC), New Delhi	SHRI S. K. KIMOTHI
Tata Engineering and Locomotive Co Ltd (TELCO), Pune	SHRI SHANTI SARUP
University College of Medical Sciences, Delhi	DR A. INDRAWAN
University of Delhi, Delhi	PROF M. C. AGRAWAL
In personal capacity (B-109, Malviya Nagar, New Delhi 110017)	PROF A. N. NANKANA
In personal capacity (20/1, Krishna Nagar, Safdarjung Enclave, New Delhi 110029)	SHRI D. R. SEN

Panel for 'Basic Methods Including Terminology, MSD 3 : 1/P-2

National Institute for Quality and Reliability (NIQR), New Delhi	SHRI G.W. DATEY (<i>Convener</i>)
Laser Science and Technology Centre, DRDO, New Delhi	DR ASHOK KUMAR
Indian Agricultural Statistics Research Institute, (IASRI), New Delhi	DR S. D. SHARMA
Indian Statistical Institute (ISI), New Delhi	PROF S. R. MOHAN
National Institute for Quality and Reliability (NIQR), New Delhi	SHRI Y. K. BHAT
Powergrid Corporation of India Ltd, New Delhi	DR S. K. AGARWAL
In personal capacity (20/1, Krishna Nagar, Safdarjung Enclave, New Delhi 110029)	SHRI D. R. SEN

Bureau of Indian Standards

BIS is a statutory institution established under the *Bureau of Indian Standards Act, 1986* to promote harmonious development of the activities of standardization, marking and quality certification of goods and attending to connected matters in the country.

Copyright

BIS has the copyright of all its publications. No part of these publications may be reproduced in any form without the prior permission in writing of BIS. This does not preclude the free use, in the course of implementing the standard, of necessary details, such as symbols and sizes, type or grade designations. Enquiries relating to copyright be addressed to the Director (Publications), BIS.

Review of Indian Standards

Amendments are issued to standards as the need arises on the basis of comments. Standards are also reviewed periodically; a standard along with amendments is reaffirmed when such review indicates that no changes are needed; if the review indicates that changes are needed, it is taken up for revision. Users of Indian Standards should ascertain that they are in possession of the latest amendments or edition by referring to the latest issue of 'BIS Catalogue' and 'Standards: Monthly Additions'.

This Indian Standard has been developed from Doc : No. MSD 3 (222).

Amendments Issued Since Publication

Amend No.	Date of Issue	Text Affected

BUREAU OF INDIAN STANDARDS

Headquarters :

Manak Bhavan, 9 Bahadur Shah Zafar Marg, New Delhi 110 002
Telephones : 2323 0131, 2323 33 75, 2323 9402

Telegrams : Manaksanstha
(Common to all offices)

Regional Offices :

	Telephone
Central : Manak Bhavan, 9 Bahadur Shah Zafar Marg NEW DELHI 110 002	{ 2323 7617 2323 3841
Eastern : 1/14 C.I.T. Scheme VII M, V. I. P. Road, Kankurgachi KOLKATA 700 054	{ 2337 8499, 2337 8561 2337 8626, 2337 9120
Northern : SCO 335-336, Sector 34-A, CHANDIGARH 160 022	{ 60 3843 60 9285
Southern : C.I.T. Campus, IV Cross Road, CHENNAI 600 113	{ 2254 1216, 2254 1442 2254 2519, 2254 2315
Western : Manakalaya, E9 MIDC, Marol, Andheri (East) MUMBAI 400 093	{ 2832 9295, 2832 7858 2832 7891, 2832 7892
Branches : AHMEDABAD. BANGALORE. BHOPAL. BHUBANESHWAR. COIMBATORE. FARIDABAD. GHAZIABAD. GUWAHATI. HYDERABAD. JAIPUR. KANPUR. LUCKNOW. NAGPUR. NALAGARH. PATNA. PUNE. RAJKOT. THIRUVANANTHAPURAM. VISAKHAPATNAM.	